NEXICA: Discovering Road Traffic Causality

Siddharth Srikanth
ssrikant@usc.edu
University of Southern California

John Krumm
jkrumm@usc.edu
University of Southern California

Jonathan Qin
qinjonat@usc.edu
University of Southern California

Abstract

Road traffic congestion is a persistent problem. Focusing resources on the causes of congestion is a potentially efficient strategy for reducing slowdowns. We present **NEXICA**, an trained binary classifier to discover which parts of the highway system tend to cause slowdowns on other parts of the highway. We use time series of road speeds as inputs to our causal discovery algorithm, and we concentrate on just the presence or absence of events (traffic slowdowns). We test on six months of road speed data from 195 different highway speed sensors in the Los Angeles area, showing that our approach is superior to state-of-the-art baselines in both accuracy and computation speed.

CCS Concepts

• **Information systems** → **Spatial-temporal systems**; **Geographic information systems**; • **Mathematics of computing** → **Maximum likelihood estimation**; • **Applied computing** → **Transportation**; • **Computing methodologies** → **Machine learning algorithms**.

Keywords

road traffic, causality, causal discovery, traffic jams

ACM Reference Format:
Siddharth Srikanth, John Krumm, and Jonathan Qin. 2025. NEXICA: Discovering Road Traffic Causality. In *The 33rd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '25), November 3–6, 2025, Minneapolis, MN, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3748636.3762738

1 Introduction

Traffic congestion cost U.S. drivers an average of 43 hours and a total of $74 billion in lost time in 2024 [14]. In 2019, the U.S. freight sector lost $74.1 billion due to slow traffic [8]. Idling engines also cause more pollution. To reduce these costs, it is important to direct resources efficiently for reducing traffic congestion.

It is important to identify the causes of the slowdowns, especially those roads that regularly cause slowdowns on other roads. These especially causal roads are good candidates for improvement. For traffic, we are fortunate to have rich data on traffic from sensors in the roads giving time series of traffic speeds measured at different locations. Causal discovery can show which locations on the roads tend to cause slowdowns at other locations.

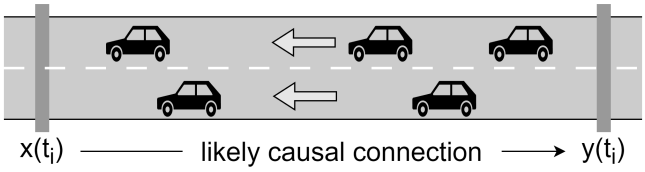


Figure 1: Road speeds at this pair of locations are likely causally connected as a slowdown at x would cause a slowdown at y after some lag in time.

The first milestone in the science of causal discovery was Granger causality, which, at its core, examines two time series, $x(t_i)$ and $y(t_i)$ [5]. We can imagine that $y(t_i)$ is the time series of road speeds at one road location, and $x(t_i)$ is the time series of road speeds at another road location, as illustrated in Figure 1. Assuming a linear, autoregressive model, Granger says that if $y(t_i)$ can be signficantly better predicted by $y(t_i)$ *and* $x(t_i)$ over predictions from $y(t_i)$ alone, then $x(t_i)$ causes $y(t_i)$.

More modern causality tests can tolerate nonlinear relationships between the variables, accommodate many more than two variables, and produce a causal graph that represents a cascade of causal relationships propagating through a network of variables. However, we found that two modern causal discovery techniques, PCMCI [13] and DYNO-TEARS [11], do not perform well for our problem. Our new algorithm, **N**atural **E**vent e**X**traction for **I**nference of **CA**usality (**NEXICA**), is novel in that it looks at binary slowdown events derived from continous-valued speed time series and it applies a machine learning classifier to the problem of causal discovery, because we have ground truth causal and noncausal pairs.

2 Related Work

In addition to traditional Granger causality, described above, there are more modern approaches to causal discovery on time series, such as PCMCI [13] and DYNOTEARS [11]. PCMCI is based on statistical tests of independence among the time series. Its goal is to build a graph of directed causal connections, where the nodes of the graph are the signals (e.g. traffic measurement stations) and the directed edges indicate a causal relationship from cause to effect. We found PCMCI to be too slow for our test data, and noted that it lacks the ability to learn from ground truth causal pairs, something that is built in to our approach.

Another category of algorithm searches directly for the entire causal graph, so-called structure learning of a directed, acyclical graph (DAG). An example is the DYNOTEARS algorithm for learning causal DAGs from time series data like ours [11]. While this work provides a foundation for us to eliminate certain edges based

Funding for this study was provided by the California Department of Transportation (Caltrans), Agreement 65A0674, Task Order 076, and NSF grant IIS-2128661. Opinions, findings, conclusions, or recommendations expressed are those of the author(s) and do not necessarily reflect the views of any sponsors, such as NSF.

SIGSPATIAL '25, Minneapolis, MN, USA

ACM ISBN 979-8-4007-2086-4/2025/11
https://doi.org/10.1145/3748636.3762738

on prior intuition, a major pitfall of this approach is that it does not support a framework to insert causal edges that we know to exist.

Previous work has focused specifically on the problem of road traffic causality. Queen and Albers present a dynamic Bayesian network for traffic forecasting that focuses on conditional independencies and causes in the relationships between measured traffic at different locations on the road network [12]. Myrovali et al. examined time series of traversal times for 10 roads in a region of Thessaloniki, Greece [10]. In the work by Jung et. al [6], the authors build an interactive system to compute and display causality connections between road locations computed with the Granger causal density test of Seth et al. [15]. Mao et al. present a state space solution for discovering traffic causality from time series of speed measurements [7].

3 Traffic Data

Our traffic data comes from California's Caltrans Performance Measurement System (PeMS) [1], which provides free downloads of traffic data from static sensors on the freeway system across all major metropolitan areas of the state. Our data was five-minute measurement intervals from "Mainline" roads in the Los Angeles area (District 7) for the first six months of 2024. The sensors are shown in Figure 2. Typical speed data for one station for one week is shown in Figure 4.

Many of the speed values in the PeMS data are imputed due to sensor dropouts [2]. The imputation methods do not necessarily preserve the anomalous traffic slowdowns on which our algorithm depends, so stations with much imputed data are not good candidates for our causality analysis. Hence, we chose 195 stations with at least 90% complete data.

3.1 Drive Time Data

For training and testing, we select ordered pairs of stations that very likely have a causal connection or very likely do not have a causal connection, as explained next in Section 3.2. Part of this ground truth selection depends on the free-flow driving time between pairs of stations. We accomplish this by constructing a drive time matrix using the Bing Maps API [9]. For each station i, we query the API to provide us with the time it takes to travel in a motorized vehicle to each other station j such that $j \neq i$. Each entry D_{ij} of this drive time matrix matrix D represents the time it takes to drive from station i to station j. It is seen trivially that $D_{ii} = 0$ for all i. Because the measured drive time from the API was dependent on traffic at that time, we selected an arbitrary time (i.e. March 3rd, 2024 at midnight). This helps ensure that congestion does not play a strong role in the computed travel time between stations, because we are most interested in the free-flow drive times for selecting ground truth pairs.

Figure 2: Locations of "Mainline" sensors in Caltrans District 7. The blue dots represent the 195 stations we used in our causality analysis, because they had sufficient speed data.

(a) Positive causal pairs. **(b) Negative causal pairs.**

Figure 3: Example ground truth causal pairs. (a): yellow shows the location of the cause (downstream from effect), and red shows the location of effect (upstream from cause). (b): both causal directions are considered as negative causal pairs.

3.2 Ground Truth

Our ultimate goal is to identify cause/effect station pairs, indicating that traffic at the "cause" station affects traffic at the "effect" station. We select these pairs from the 195 measurement stations that had enough non-imputed data as described above. In selecting ground truth pairs, we consider all candidate causal pairs (i, j, k), where i is the potential causal station, j is the potential affected station, and k is the temporal lag from cause to effect. The lag k is an integer multiple of five minutes, because the time series is sampled every five minutes. This set contains ${}^{N}P_2 l_{max} = N(N-1)l_{max}$ unique edges, where l_{max} represents the total number of lags considered and $N = 195$ is the total number of stations considered.

From this set of candidate causal pairs, we select ground truth pairs as positive (causal) or negative based on several criteria. We attempt to choose pairs where the causality is very clear, and thus we are conservative in our selections in order to create an accurate ground truth set. We are essentially looking for obvious causal pairs like the one illustrated in Figure 1. Our criteria for selecting positive ground truth pairs are: **(1) Distance required to travel**, classifying edges as negative if their drive time is too long, based on a maximum congestion propagation speed [4]; (2) **Road Number and Direction**, assuming majority of traffic events are contained to a specific road and direction (i.e. CA-110 North or I-105 East); (3) **Direction of Travel**, i.e. northbound or southbound on the **same** highway; (4) **Propagation Direction**, ensuring traffic only propagates upstream.

All edges that meet these criteria are labeled positive in our dataset. For example, in Figure 3a, we see nearby stations where the causal station i is upstream from the affected station j, which is correspondingly labeled as a positive ground truth pair in our dataset. It is important when using the driving time between two stations to construct the ground truth, we use the drive time from the affected station j to the causal station i. This gives us context as to how long it would theoretically take us to see traffic from i appear at j. The remaining edges that do not meet this criteria are placed in a pool of possible negative edges, meaning they may not have a causal connection. For our experiments, we consider the entire ground truth data set along with a "balanced" dataset that is constructed by sorting these edges in descending order by the

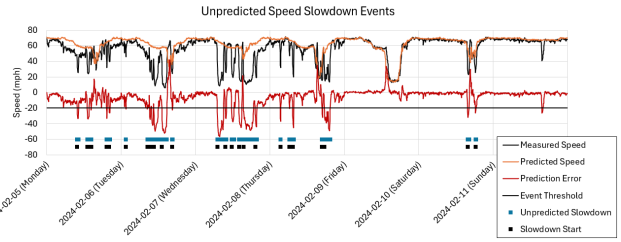
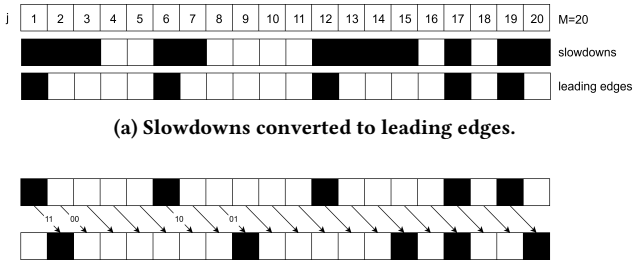


Figure 4: Traffic slowdown events are the leading edges of unpredicted slowdowns. This plot shows one week of speed data from one measurement station and its associated slowdown events.



(a) Slowdowns converted to leading edges.

(b) Lag of one, counting different kinds of correspondences between two sequences of slowdown leading edges.

Figure 5: (a) and (b) show slowdowns as black squares.

driving time between them, D_{ij}, and then taking an equal number of negative edges from this sorted list of edges. In the balanced ground truth set, the minimum drive time of any negative edge is over 75 minutes, meaning we can be fairly certain that none of the negative edges represent a causal connection. Some of these negative ground truth pairs are shown in Figure 3b.

Using these criteria, and a set of 195 stations with $l_{max} = 8$, we consider $195 * 194 * 8 = 302,640$ total candidate edge/lag tuples, from which 1771 as labeled positive ground truth edges.

4 Event-Based Traffic Causality

Our algorithm for inferring causaility is based on the temporal correspondence of unexpected, discrete traffic slowdown events. The speed measurement of station i at time j is $s_i(t_j)$ for $i \in [1, N]$ and $j \in [1, M]$. Here $N = 195$ measurement stations and $M = 52,416$ five-minute intervals in the first six months of 2024. As described above, each station i reports the average speed over all lanes every five minutes, thus t_j represents a discrete time that is a whole multiple of five minutes. To find unexpected slowdown events at a station, we must have a notion of normal traffic speed at the station. For each measurement station, we compute the median speed for each time slot in a generic week as its speed prediction. For example, the predicted speed at 10 am on a Monday is the median of all the speeds observed at that station at 10 am on all Mondays. Figure 4 shows one week of traffic speed data for one station. The black curve shows the measured speeds, and the orange curve shows the median week computed over all six months of data. This method tends to capture the normal variations in speed, ignoring unusual variations by using the median.

The predicted traffic speed is $\hat{s_i}(t_j)$, and the prediction error is $s_i(t_j) - \hat{s_i}(t_j)$. This is shown as the red curve in Figure 4. The prediction error is negative when the actual traffic speed is slower than the prediction, indicating an unexpected slowdown. We declare a traffic slowdown when the prediction error fraction is less than a threshold α, i.e. when $\frac{s_i(t_j) - \hat{s_i}(t_j)}{\hat{s_i}(t_j)} < -\alpha$. We used $\alpha = 0.25$.
An example of the inferred slowdowns is shown as blue squares in Figure 4, and a close-up example is shown as the top signal in Figure 5a. We are interested in sudden speed disruptions, so we concentrate on the leading edges of traffic slowdowns indicating the start of a slowdown as shown as the bottom signal in Figure 5a.

To infer causality between two stations at a given lag, we use features that describe the correspondence of traffic slowdown events on a lagged version of the candidate cause and the candidate effect, as illustrated in Figure 5b. There are four possible scenarios for a single corresponding pair of event time slots for the cause and effect sequence. If zero (0) represents no event and one (1) represents an event, then the four possible pairs are 00, 01, 10, and 11, and we notate the four corresponding counts as A_{00}, A_{01}, A_{10}, and A_{11}. With some pairs serving as ground truth, we train a forest of decision trees using these four featues to create a model to infer the causality of new candidate pairs. We also have an alternate approach that uses the four features to compute the probability of causality directly [16].

Framework	Balanced Scalar Threshold AUC	Balanced RF AUC	Full RF AUC	Time
DYNOTEARS	0.5769	0.5626	0.4939	111:32:06
F-PCMCI*	0.5003	0.5000	0.4999	33:44:12
Granger	**0.7156**	0.9185	0.7119	00:22:54
NEXICA	-	**0.9995**	**0.8851**	**00:01:51**

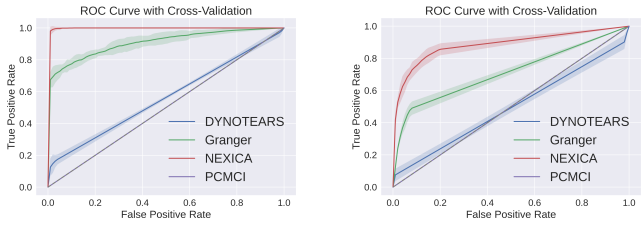
Table 1: Area Under Curve (AUC) of NEXICA versus existing baselines on ground truth data. Balanced Scalar Threshold AUC uses the best feature (highest AUC) from each experiment as a scalar threshold, whereas Balanced RF AUC uses all the parameters learned from the experiment. Full RF AUC uses the entire ground truth dataset. Time is in *HH:MM:SS*.

5 Results

We evaluated our approach on the first six months of PeMS data from 2024 on 195 select traffic sensors with at least 90% complete data. In preliminary experiments, we tested our approach using data without the restriction of having a high threshold of data completeness. We found that our model struggled to identify anomalous periods and did not accurately classify our ground-truth data. We attribute this to the event-based nature of our approach: imputed data is by nature non-anomalous and, therefore, hides events.

We note that the size of our experiment is large compared to previous work in traffic causality discovery. For example, [12] looked at one road intersection, and [7] looked at two small clusters of road sensors.

Because the number of positive ground truth edges in our dataset is significantly less than the number of negative edges (since most



(a) Balanced Random Forest ROC Curves.

(b) Full Random Forest ROC Curves.

Figure 6: ROC curves for various approaches on both the balanced and full datasets against SoTA, with AUC ± standard deviation of 2 from cross validation. Data and timing is presented in Table 1. For traffic causality, NEXICA outperforms existing methods in causality inference.

station pairs and station lags are causally improbable/impossible), we not only evaluate NEXICA on the full ground truth data, but we also evaluate on a so-called "balanced" set of ground truth data. We construct our balanced set by reducing the number of negative edges to only consider the most extreme examples, i.e. edges with the highest driving times D_{ij}. This way, we retain a $1:1$ ratio between positive and negative ground truth edges in our dataset to ensure a more realistic evaluation of our results.

Table 1 shows the performance of our algorithm versus state-of-the-art benchmarks PCMCI and DYNOTEARS described in Section 2, all run on thte same hardware. For the PCMCI baseline, we found it computationally expensive to run using the entire time-series, so we ran it with only 1 week of timeseries data using the same 195 stations. Furthermore, we ran an improved version of the PCMCI algorithm, F-PCMCI [3] for faster and more robust inference.

We learn a complex decision boundary for each candidate method by using a random forest of decision trees, each of which using 1000 estimators and leveraging the ground truth data as labels during training and testing. We then extract a classification probability p_{forest} for each candidate tuple (i, j, k), determined from the classification vote in the ensemble. We train and test each forest on the balanced and full ground truth data using 5-fold cross-validation.

For decision tree features, we we use the correspondence counts A_{00}, A_{01}, A_{10}, and A_{11} for our method (not distance nor any other features that we used to select ground truth). Our baselines use the following computed features: DYNOTEARS computes the weight of the edge in the final DAG; F-PCMCI computes the weight of the edge in the final graph along with its significance (p-value); Granger causality computes the values of a Params F test, SSR F test, SSR Chi Squared test, and a Likelihood-ratio test.

We present the results of NEXICA versus the existing state-of-the-art baselines in Table 1. Since thresholding on a single correspondence count is not well-defined within our framework, Granger causality outperforms our method in the Balanced Scalar Threshold Evaluation. However, using all four correspondence counts (A_{00} etc.) with decision trees enables NEXICA to outperform all state-of-the-art baselines in both the balanced and full sets. Our best classifier achieved near-perfect accuracy when evaluated with cross-validation.

In Figure 6, we plot the ROC curves of all 3 methods from Table 1. The visually peculiar behavior from the baselines DYNOTEARS and F-PCMCI are because these approaches yield a nonzero weight for a very small number of edges in the causal graph. As a result, the binary classifier struggles to learn an appropriate model from these algorithms' features.

6 Conclusion

Our goal is to discover which locations in the road network tend to cause slowdowns at other locations. These especially causal locations are good candidates for further investigation and investment. Our new approach, NEXICA, uses natural traffic slowdown events detected from traffic speed time series along with a maximum likelihood algorithm that computes the probability of one sequence of events causing another. Our experiments show that it is both fast and accurate compared to the state-of-the-art approaches. For future work, we envision expanding to other regions and road types, and improving the core mathematical algorithm to account for causal graphs with multiple causal edges and signal nodes.

References

[1] California Department of Transportation. 2023. Caltrans Performance Measurement System. https://pems.dot.ca.gov/ Accessed: 2025-02-03.
[2] Caltrans. 2020. *PeMS Introduction and User Guide.* https://pems.dot.ca.gov/Papers/PeMS_Intro_User_Guide_v6.pdf
[3] Luca Castri, Sariah Mghames, Marc Hanheide, and Nicola Bellotto. 2023. Enhancing Causal Discovery from Robot Sensor Data in Dynamic Scenarios. In *Conference on Causal Learning and Reasoning (CLeaR).*
[4] Wen-peng Fei, Guo-hua Song, Fan Zhang, Yong Gao, and Lei Yu. 2017. Practical approach to determining traffic congestion propagation boundary due to traffic incidents. *Journal of Central South University* 24, 2 (2017), 413–422.
[5] C. W. J. Granger. 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37, 3 (1969), 424–438.
[6] Chanyoung Jung, Soobin Yim, Giwoong Park, Simon Oh, and Yun Jang. 2024. CATOM: Causal Topology Map for Spatiotemporal Traffic Analysis with Granger Causality in Urban Areas. *IEEE Transactions on Visualization and Computer Graphics* (2024).
[7] Jiannan Mao, Hao Huang, Yu Gu, Weike Lu, Tianli Tang, and Fan Ding. 2025. A convergent cross-mapping approach for unveiling congestion spatial causality in urban traffic networks. *Computer-Aided Civil and Infrastructure Engineering* 40, 3 (2025), 301–322.
[8] Niall McCarthy. 2020. Chart: Congestion Costs U.S. Cities Billions Every Year. https://www.statista.com/chart/21085/annual-economic-losses-from-traffic-congestion/ Accessed: 2025-02-18.
[9] Microsoft. 2024. Bing Maps API. https://www.bingmapsportal.com/ Accessed: 2024-06-01.
[10] Glykeria Myrovali, Theodoros Karakasidis, Georgia Ayfantopoulou, and Maria Morfoulaki. 2021. Spatio-Temporal Causal Relations at Urban Road Networks; Granger Causality Based Networks as an Insight to Urban Traffic Dynamics. In *Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 1.* Springer, 791–804.
[11] Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. 2020. Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics.* Pmlr, 1595–1605.
[12] Catriona M Queen and Casper J Albers. 2009. Intervention and causality: forecasting traffic flows using a dynamic Bayesian network. *J. Amer. Statist. Assoc.* 104, 486 (2009), 669–681.
[13] Jakob Runge. 2020. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence.* Pmlr, 1388–1397.
[14] Monica Sager. 2024. You Wasted $771 Sitting in Traffic Last Year. https://www.newsweek.com/traffic-study-wasted-money-economy-congestion-2012671 Accessed: 2025-02-18.
[15] Anil K Seth. 2005. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems* 16, 1 (2005), 35–54.
[16] Siddharth Srikanth, John Krumm, and Jonathan Qin. 2025. NEXICA: Discovering Road Traffic Causality (Extended arXiv Version). arXiv:2508.09447 [cs.LG] Extended version of paper in ACM SIGSPATIAL 2025.